IN THE CLAIMS

1. (Currently amended) A method of using a computer processor to improve speech recognition performance in an audio-visual speech recognition system comprising the steps of:

receiving ~~at least one of~~ audio data and visual data associated with an input spoken utterance;

using the computer processor to select between an acoustic-only data model and an acoustic-visual data model based on a <u>level of degradation of the visual data</u> ~~condition associated with a visual environment~~; and

using the computer processor to decode at least a portion of ~~the~~ at least one of <u>the</u> audio data and <u>the</u> visual data associated with the input spoken utterance using the selected data model.

2. (Original) The method of claim 1, further comprising the step of storing the acoustic-only data model and the acoustic-visual data model in memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located.

3. (Original) The method of claim 1, wherein the model selection step is based on a likelihood ratio test.

4. (Original) The method of claim 3, wherein the model selection step further comprises selecting the acoustic-only data model when a result of the likelihood test is not greater than a threshold value.

5. (Original) The method of claim 3, wherein the model selection step further comprises selecting the acoustic-visual data model when a result of the likelihood test is not less than a threshold value.

6. (Original) The method of claim 5, wherein the threshold value is based on a cost associated with a recognition error.

2

7. (Original) The method of claim 3, wherein the likelihood ratio test is based on one or more observations of a given visual feature.

8. (Original) The method of claim 7, wherein the given visual feature is associated with the mouth region of a speaker of the input utterance.

9. (Original) The method of claim 1, wherein model selection is performed at a rate substantially equivalent to an observation rate associated with the audio-visual speech recognition system.

10. (Currently amended) Apparatus to improve speech recognition performance in an audio-visual speech recognition system the apparatus comprising:

a memory; and

at least one processor coupled to the memory and operative to: (i) receive ~~at least one of~~ audio data and visual data associated with an input spoken utterance; (ii) select between an acoustic-only data model and an acoustic-visual data model based on a <u>level of degradation of the visual data</u> ~~condition associated with a visual environment~~; and (iii) decode at least a portion of ~~the~~ at least one of <u>the</u> audio data and <u>the</u> visual data associated with the input spoken utterance using the selected data model.

11. (Original) The apparatus of claim 10, wherein the acoustic-only data model and the acoustic-visual data model are stored in the memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located.

12. (Original) The apparatus of claim 10, wherein the model selection operation is based on a likelihood ratio test.

13. (Original) The apparatus of claim 12, wherein the model selection operation further

comprises selecting the acoustic-only data model when a result of the likelihood test is not greater than a threshold value.

14. (Original) The apparatus of claim 12, wherein the model selection operation further comprises selecting the acoustic-visual data model when a result of the likelihood test is not less than a threshold value.

15. (Original) The apparatus of claim 14, wherein the threshold value is based on a cost associated with a recognition error.

16. (Original) The apparatus of claim 12, wherein the likelihood ratio test is based on one or more observations of a given visual feature.

17. (Original) The apparatus of claim 16, wherein the given visual feature is associated with the mouth region of a speaker of the input utterance.

18. (Original) The apparatus of claim 10, wherein model selection is performed at a rate substantially equivalent to an observation rate associated with the audio-visual speech recognition system.

19. (Currently amended) An article of manufacture for use with a computer processor to improve speech recognition performance in an audio-visual speech recognition system, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

receiving ~~at least one of~~ audio data and visual data associated with an input spoken utterance;

using the computer processor to select between an acoustic-only data model and an acoustic-visual data model based on a <u>level of degradation of the visual data</u> ~~condition associated with a visual environment~~; and

4

using the computer processor to decode at least a portion of ~~the~~ at least one of <u>the</u> audio data and <u>the</u> visual data associated with the an input spoken utterance using the selected data model.

20. (Original) The article of claim 19, further comprising the step of storing the acoustic-only data model and the acoustic-visual data model in memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located.

21. (Currently amended) An audio-visual speech recognition system, comprising:

a memory; and

at least one processor coupled to the memory and operative to: (i) receive ~~at least one of~~ audio data and visual data associated with an input spoken utterance; (ii) select between an acoustic-only data model and an acoustic-visual data model based on a <u>level of degradation of the visual data</u> ~~condition associated with a visual environment~~; and (iii) decode at least a portion of ~~the~~ at least one of <u>the</u> audio data and <u>the</u> visual data associated with the input spoken utterance using the selected data model, wherein the acoustic-only data model and the acoustic-visual data model are stored in the memory such that model selection is made by shifting one or more pointers to one or more memory locations where the selected model is located.

22. (Currently amended) A method of using a computer processor to improve speech recognition performance in a speech recognition system comprising the steps of:

receiving one or more frames of ~~at least one of~~ audio data and visual data associated with an input spoken utterance;

using the computer processor to select for a given frame between a first data model and at least a second data model based on a <u>level of degradation of the visual data</u> ~~given condition~~; and

using the computer processor to decode at least a portion of ~~the~~ at least one of <u>the</u> audio data and <u>the</u> visual data associated with the input spoken utterance for the given frame using the selected data model.